

A System for Rapidly and Accurately Collecting Patients' Race and Ethnicity

David W. Baker, MD, MPH, Kenzie A. Cameron, PhD, Joseph Feinglass, PhD, Jason A. Thompson, BA, Patricia Georgas, BA, Shawn Foster, BA, Deborah Pierce, BA, and Romana Hasnain-Wynia, PhD

People of color and racial/ethnic minorities in the United States often receive lower-quality health care than Whites.^{1,2} The first step toward addressing this problem is for health care providers to routinely collect data on patients' race, ethnicity, and language and link these data to measures of quality, safety, and utilization.^{1–5} With such information, provider organizations can target quality-improvement programs to reduce disparities at their own institutions.

A recent survey found that only 78% of US hospitals systematically collected data on patients' race or ethnicity.⁶ Among those that did, more than half relied on registration clerks' impressions of patients' race and ethnicity rather than directly asking patients. When compared with patients' self-report, staff impressions of patients' race/ethnicity are reasonably accurate for Whites and Blacks but much less accurate for other groups.^{7–11} In addition, relying on staff impressions results in high rates of patients having missing data or being classified as "unknown" or "other."^{4,7} Because of this, expert panels have recommended that patients' race and ethnicity should be collected by self-report.^{1–5}

However, even if there is widespread adoption of the recommendation to collect patients' race/ethnicity by self-report, current data collection methods are suboptimal. Providers typically ask questions that comply with the Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity from the Office of Management and Budget (OMB) and are used by the US Census Bureau: (1) "Do you consider yourself Latino or Hispanic?" and (2) "Which category best describes your race: White, Black or African American, Asian, Native Hawaiian or Pacific Islander, American Indian or Alaska Native, multiracial, or another race I did not mention?"¹² More-detailed information is needed to address health and health care

Objectives. We assessed the feasibility of collecting race/ethnicity data from patients using their own preferred racial/ethnic terms.

Methods. The 424 patients described their race/ethnicity using their own categories, and we compared their descriptions with their responses to the questions (1) "Do you consider yourself Latino or Hispanic?" and (2) "Which category best describes your race?" (7 response options in our computer interview). We also determined patients' preferences between the 2 approaches.

Results. The proportions of patients who described themselves with 1, 2, 3, or 4 terms were 46%, 33%, 14%, and 6%, respectively; 2 said only "American" (1%), and 1 refused to answer (0.5%). The average completion time was 37 ± 17 seconds. Rates of missing values and categorization as "other" race were lower than with the closed questions. Agreement between racial/ethnic categorization with open-ended and closed responses was 93% ($\kappa=0.88$). Latino/Hispanic and multiracial/multiethnic individuals were more likely to prefer using their own categories to describe their race/ethnicity.

Conclusions. Collecting race/ethnicity data using patients' own racial/ethnic categories is feasible with the use of computerized systems to capture verbatim responses and results in lower rates of missing and unusable data than do standard questions. (*Am J Public Health.* 2006;96:532–537. doi:10.2105/AJPH.2005.062620)

disparities.⁵ For example, several studies have emphasized the heterogeneity within the Hispanic/Latino population^{13–18}; categorizing individuals of Mexican, Cuban, and Puerto Rican ancestry as simply Hispanic/Latino may obscure disparities in access, utilization, or quality of care.¹⁹ Important differences in risk factors for cardiovascular disease have been reported across Asian subgroups.^{20,21}

The increasing heterogeneity of the population of the United States makes the deficiencies of the traditional OMB race/ethnicity questions even more salient. How should immigrants from Poland, Iraq, and India classify themselves? Classifying them as White blurs important distinctions in their health beliefs, behaviors, and treatment by health care providers. Alternatively, classifying people from Iraq and India as Asian combines them with individuals of Hmong or Japanese ancestry, which is equally inappropriate. The OMB standards were expressly intended as a minimum data requirement, but few health care

providers ask more detailed questions. In part these questions are not asked because collecting more-detailed information is problematic if a clerk must read a long list of options from which patients are asked to choose. An alternative is to allow patients to self-describe themselves and record their verbatim responses with computerized data collection tools. This should allow providers to analyze their data at different levels of complexity, ranging from small-subgroup analyses using unique categories (e.g., Korean or Puerto Rican) up to more coarse, aggregated categories (e.g., Asian or Hispanic/Latino).

We undertook this project to develop an instrument that could rapidly and accurately collect patients' self-described race/ethnicity. The first phase tested the feasibility of allowing patients to self-identify themselves with whatever terms they wanted. The second phase tested a computerized tool for capturing this information and sought to examine the terms used by patients to self-identify

themselves, rates of missing values and refusals, and time required for data collection.

METHODS

Study Site and Population

Our study was conducted in the General Internal Medicine Clinic of the Northwestern Medical Faculty Foundation in Chicago, Ill, from December 2003 to April 2004. Patients were eligible if they were aged 18 years and spoke English fluently. We approached patients after their visit and asked them to participate, stating: "The purpose of this research study is to understand what patients think about being asked to give information about their race or ethnic background." A parking pass was offered as an incentive, and we did not offer additional incentives to people who refused. After finishing the interview, a research assistant approached the next patient coming to the discharge area. Thus, the study population was a convenience sample and not a random sample. Patients who refused were asked their age, and the research assistant recorded the patients' gender and probable race/ethnicity.

Initial Instrument Development and Testing (Phase 1)

Participating patients were first asked to describe their race/ethnicity with any terms they wanted to use. After each response, participants were asked, "Anything else?" Up to 4 terms were recorded verbatim. We then explained that we wanted to ask the race/ethnicity questions using a different method and asked questions based on those used by the OMB: (1) "Do you consider yourself Latino or Hispanic?" and (2) "Which category best describes your race: White, Black or African American, Asian, Native Hawaiian or Pacific Islander, American Indian or Alaska Native, multiracial, or another race I did not mention?" We then asked which of the 2 methods of obtaining race/ethnicity information the patient preferred and the strength of the preference (somewhat or strongly prefer).

Next, we asked patients how comfortable they would feel stating their race/ethnicity if asked to do so by a clerk at a hospital or clinic. We tested 4 statements to determine if they increased patients' level of comfort when

providing their race/ethnicity. These results have been described in a previous paper²² and are not presented here. The preferred statement, for both Whites and non-Whites, was "We want to make sure that all our patients get the best care possible, regardless of their race or ethnic background. We would like you to tell us your race or ethnic background so that we can review the treatment that all patients receive and make sure that everyone gets the highest quality of care." A modified version of this statement was, therefore, chosen as the introduction to race/ethnicity question in phase 2.

Testing of the Final Instrument (Phase 2)

A computerized instrument for collecting patients' responses was developed on the basis of the pilot test. A list of terms patients might use to describe their race/ethnicity was developed and used in the data collection tool as a look-up table with a drop-down menu. Every term was given a unique code to allow us to aggregate individual terms into groups that approximated the OMB categories.

During the interview, patients were first read this statement: "We want to make sure that all our patients get the best care possible, regardless of their race or ethnic background. We would like you to tell us your race or ethnic background so that we can review the treatment that all patients receive and make sure that everyone gets the highest quality of care. You can use general terms such as White, African American, Latino, or Asian or you can use more specific terms like Irish, Mexican, Hawaiian, or Navajo. You can use more than 1 term if you like." The results of pilot testing suggested that asking "Anything else?" after each response led some individuals to search for answers and give redundant responses (e.g., White and "Caucasian"). We therefore simply allowed respondents time to say other terms, but we did not solicit additional responses. The database had drop-down windows with terms listed alphabetically so that the research assistant only had to type the first few letters of what the patient said and then select the desired term by double-clicking on it to enter it into the database. The database automatically proceeded to the next field after entry to allow participants to say additional terms. Time stamps were

automatically recorded during this process to measure the time required for completion.

Data Analysis

Data analysis was performed with Stata version 8.0 (Stata Corp, College Station, Tex). Descriptive statistics were determined for all questions. Each unique race and ethnicity term used by patients to describe themselves was assigned to 1 of the following race/ethnicity categories (see the footnotes for Table 2): White, Black or African American, Hispanic or Latino, Asian, Native Hawaiian or Pacific Islander, or American Indian/Alaska Native. Individuals who reported a single race or ethnicity or who used 2 or more terms that mapped to a single race/ethnicity category (e.g., German and Polish; Cuban and Puerto Rican) were classified as belonging to 1 of the 6 unique categories previously mentioned. Individuals who used terms from 2 or more categories or who described themselves as multiracial were classified as "multiracial."

Differences in patients' preferences for describing their race/ethnicity with their own terms compared with choosing from a list of options were assessed with pairwise χ^2 tests. We then used multinomial logistic regression to examine racial/ethnic differences in patients' preferences with adjustment for age, gender, and study phase (1 or 2); patients who thought the 2 methods were "about the same" were used as the reference group in the model. One Native American patient, 2 patients who described themselves only as "American," and 2 patients who refused to answer the open-ended questions were excluded from the final multivariate model. Two-tailed tests were used for all analyses, and a final *P* value of .05 was used to determine statistical significance.

Finally, we analyzed differences between patients' self-reported descriptions of their race/ethnicity and their responses to the OMB questions regarding Hispanic/Latino ethnicity and race. We used patients' own terms to categorize them into groups that approximated the categories used in federal studies, as described earlier. Because the open-ended self-report method does not ask separate questions for Hispanic ethnicity and race, it was not possible to classify all individuals as non-Hispanic White, Hispanic White, non-Hispanic Black, Hispanic Black, and so forth. Therefore, to

TABLE 1—Characteristics of Study Participants: Chicago, Ill, December 2003 to April 2004

	Phase 1 (n = 220)	Phase 2 (n = 204)
Female, %	67	73
Age, mean years \pm SD	44 \pm 15	46 \pm 6
Duration of care at study site, %		
\leq 1 y	19	27
2–5 y	36	32
6–10 y	21	18
>10 y	24	23
Number of visits in last year, %		
0–1	15	8
2–3	30	27
4–7	29	29
\geq 8	27	36

assess the agreement between the 2 systems, we compared the Latino/Hispanic category from the open-ended questions with the Hispanic White category derived from the OMB questions; the 3 participants who were classified as Black Hispanic on the basis of the OMB questions were considered to not have a matching category on the basis of the classification from self-report (i.e., 100% disagreement between the 2 methods). The percentage agreement between the 2 methods of collecting race/ethnicity data was calculated along with the κ statistic²³ to determine the agreement adjusted for chance.

RESULTS

Participant Characteristics

Of the 373 patients who were approached and asked to participate in the initial instrument development (phase 1), 220 (59%) gave informed consent and completed the interview. The mean age of participants was 44 years, and 67% were female (Table 1). Almost all patients had been cared for in the clinic for more than 1 year and had 2 or more visits over the previous year. For the testing of the final instrument (phase 2), 367 patients were approached and asked to participate, and 204 (56%) participated. The characteristics of phase 2 participants were similar to those for phase 1 (Table 1). The proportion

of women and the estimated age were similar for participants and nonparticipants.

Responses to Open-Ended Question About Race/Ethnicity

When participants in the initial testing (phase 1) were asked to describe their race or ethnicity with any terms they wanted, 46% of participants described themselves with a single term, 33% used 2 terms, 14% used 3 terms, and 6% used 4 terms; 2 patients (1%) said only “American,” and 1 patient (0.5%) refused to answer. Among those who specified 2 or more terms, many used redundant terms that mapped to a single composite race/ethnicity category (e.g., White and “Caucasian,” or Black and African American). The distribution of participants’ responses is shown in Table 2. During phase 2, 77% of participants described themselves with a single term, 17% used 2 terms, 4% used 3 terms, and 2% used 4 terms ($P < .001$ for χ^2 test comparing the number of terms used in phase 1 vs phase 2); no patients in phase 2 used the term “American,” and no patients refused to answer.

Responses to the OMB Questions

Of the 424 participants, 44 (10%) responded yes to the question “Are you Latino or Hispanic?”; 4 participants refused to answer the question, including 1 individual who described himself as a Spaniard. For the close-ended race question, 212 (50%) chose White, 128 (30%) Black or African American, 20 (5%) Asian, 2 (0.5%) Pacific Islander or Native Hawaiian, and 3 (0.7%) American Indian or Alaska Native. In addition, 13 (3%) chose “multiracial,” 35 (8%) chose Other, and 11 (3%) refused to answer. Latinos were much more likely than Whites to choose Other as their race (47% vs 4%; $P < .001$) or to refuse to answer (14% vs 1%; $P < .001$).

Preferences for Using Own Terms vs Choosing from a List

Overall, 23.8% strongly preferred using their own words to describe their race/ethnicity, 8.0% somewhat preferred using their own words, 41.3% had no preference, 11.8% somewhat preferred choosing from a list, and 15.1% strongly preferred choosing from a list (Table 3). However, preferences varied among Whites, African Americans,

TABLE 2—Self-Descriptions of Race/Ethnicity

	Phase 1 (n = 220), No. (%)	Phase 2 (n = 204), No. (%)
White or “Caucasian” only	23 (10.5)	66 (32.4)
White or “Caucasian” and specific terms ^a	40 (18.2)	28 (13.7)
Specific terms only ^a	25 (11.4)	11 (5.4)
African American or Black only	66 (30.0)	61 (29.9)
African American or Black and specific terms ^b	4 (1.8)	1 (0.5)
Specific terms only ^b	3 (1.4)	0 (0.0)
Latino or Hispanic only	7 (3.2)	4 (2.0)
Latino or Hispanic and specific terms ^c	4 (1.8)	0 (0.0)
Specific terms only ^c	6 (2.7)	11 (5.4)
Asian only	2 (0.9)	3 (1.5)
Asian and specific terms ^d	1 (0.5)	1 (0.5)
Specific terms only ^d	6 (2.7)	4 (2.0)
American Indian or Native American only	2 (0.9)	0 (0.0)
Pacific Islander, Native Hawaiian, or specific terms	0 (0.0)	0 (0.0)
Multiracial	8 (3.6)	0 (0.0)
Used 2 or more descriptors from above categories	18 (8.2)	14 (6.9)
“American”	2 (0.9)	0 (0.0)
Other	1 (0.5)	0 (0.0)
Refused	2 (0.9)	0 (0.0)

^aIncluding Anglo, Arab, Barbadian, Basque, British/British Island, Czech, Danish, Eastern European, English, European, French, German, Greek, Hungarian, Iranian, Irish, Italian, Jewish, Lebanese, Mediterranean, Middle Eastern, Middle European, North American, Northern European, Norwegian, Palestinian, Pennsylvania Dutch, Polish, Russian, Scandinavian, Scottish, Serbian, Slovak, Swedish, Ukrainian, WASP, Western European

^bIncluding African, Haitian, Jamaican, Person of Color, Soul Sister No. 1, West African

^cIncluding Brazilian, Chicano, Chilean, Colombian, Cuban, Mexican, Puerto Rican, Salvadoran, Spanish

^dIncluding Chinese, Filipino, Indian, Japanese, Korean, Vietnamese

Latinos/Hispanics, and multiracial/multiethnic individuals, with non-Whites being more likely to prefer using their own words ($P = .002$, $P = .02$, and $P = .003$ for pairwise comparisons between each of the groups vs Whites). For example, the proportion of participants who somewhat or strongly preferred to use their own words was 24%, 35%, 46%, and 52%

TABLE 3—Preferences for Using Own Words vs Choosing From a List, Stratified by Self-Reported Race/Ethnicity Categories.

	Preference for Using Own Words vs Choosing From List				
	White or "Caucasian," No. (%)	Black or African American, No. (%) ^a	Latino or Hispanic, No. (%) ^b	Multiracial, No. (%) ^c	Total, No. (%) ^d
Much prefer own words	31 (16.0)	37 (27.2)	12 (32.4)	16 (48.5)	101 (23.8)
Somewhat prefer own words	15 (7.7)	10 (7.4)	5 (13.5)	1 (3.0)	34 (8.0)
About the same	83 (42.8)	66 (48.5)	9 (24.3)	7 (21.2)	175 (41.3)
Somewhat prefer choosing from list	27 (13.9)	11 (8.1)	6 (16.2)	4 (12.1)	50 (11.8)
Much prefer choosing from list	38 (19.6)	12 (8.8)	5 (13.5)	5 (15.2)	64 (15.1)

^a*P* = .007 compared with White/Caucasian.^b*P* = .058 compared with White/Caucasian.^c*P* = .001 compared with White/Caucasian.^dRespondents who described themselves as Asian (*n* = 19), American Indian/Alaska Native (*n* = 1), or "American" (*n* = 2), as well as those who refused (*n* = 2), are not shown as separate columns but are included in totals.

among Whites, African Americans, Latinos/Hispanics, and multiracial/multiethnic individuals, respectively (Table 3). In multivariate analysis with adjustment for age, gender, and study phase, the adjusted odds ratios for preferring to use one's own words were 1.1 (95% confidence interval [CI]=0.7, 1.9) for African Americans, 3.0 (95% CI=1.2, 7.5) for Latinos/Hispanics, and 3.2 (95% CI=1.2, 8.6) for multiracial/multiethnic individuals (all vs Whites). Conversely, African Americans were less likely than Whites to say they preferred to choose from a list (adjusted odds ratio=0.4;

95% CI=0.2, 0.8); no differences were seen for Latinos/Hispanics and multiracial/multiethnic individuals compared with Whites. Age and gender were not related to preferences, but patients in phase 2 were more likely to say they thought that describing their race/ethnicity in their own words and choosing from a list were "about the same."

Categorization With Patients' Own Terms vs Choosing from OMB Categories

There was excellent agreement between the categorization based on patients' own

words compared with their responses to their choices for the questions on Hispanic ethnicity and race (Table 4). After excluding subjects who chose Other or refused to answer 1 or both of the OMB questions, there was 93% agreement between the categories created with patients own terms and the categories based on the OMB questions (κ statistic=0.88).

Despite the overall high level of agreement between the 2 classification systems, the areas of disagreement were noteworthy and important. Among the 37 patients who described themselves as Latino, Hispanic, or being from a country in Central or South America, 17 (46.0%) chose Other as their race when asked to choose from a list. Another 8 (21.6%) patients refused to choose any race term. In addition, a "Spaniard from Spain" refused to answer the question about Hispanic ethnicity.

Rates of choosing the Other race option from the OMB list or refusing to answer the race question were lower among self-described Whites and African Americans but still problematic (3.1% and 8.3%, respectively). For example, 1 patient self-described herself as Jamaican and Barbadian (i.e., from Barbados) and chose Other as her race; a patient from "West Africa" did the same. Of the 6 patients who described themselves as being of Middle Eastern origin (i.e., Arab, Iranian, Palestinian,

TABLE 4—Comparison of Categorization into Major Racial/Ethnic Groups by OMB Questions and Open-Ended Questions.

Categorization Based on OMB Questions	Categorization Based on Responses to Open-Ended Questions							
	White (<i>n</i> = 194), No. (%)	Black or African American (<i>n</i> = 133), No. (%)	Latino or Hispanic (<i>n</i> = 37), No. (%)	Native Asian (<i>n</i> = 19), No. (%)	Native American or Alaska Native (<i>n</i> = 1), No. (%)	Hawaiian or Pacific Islander (<i>n</i> = 0), No. (%)	Multiracial (<i>n</i> = 33), No. (%)	Other/Missing (<i>n</i> = 33), No. (%) ^a
Non-Hispanic White	187 (96.4)	0	1 (2.7)	0	0	0	10 (30.3)	2 (50.0)
Non-Hispanic Black	0	119 (89.5)	0	0	0	0	5 (15.2)	0
Hispanic White	0	0	10 (27.0)	0	0	0	0	1 (25.0)
Asian	1 (0.5)	0	0	18 (94.7)	0	0	1 (3.0)	0
Native American or Alaska Native	0	0	0	1 (5.3)	1 (100.0)	0	0	1 (25.0)
Native Hawaiian or Pacific Islander	0	0	0	0	0	0	2 (6.1)	0
Multiracial	0	3 (2.3)	1 (2.7)	0	0	0	9 (27.3)	0
Other	5 (2.6)	7 (5.3)	17 (46.0)	0	0	0	6 (18.2)	0
Missing	1 (0.5)	4 (3.0)	8 (21.6)	0	0	0	0	0

Note. OMB = Office of Management and Budget. Respondents who were classified as Hispanic Black by the OMB questions (*n* = 3) were not included in this table, because no analogous open-ended category was used.

^aIncludes 2 participants who responded "American" and 2 who refused to answer.

or “Middle Eastern”), 2 chose their race as Other and 1 refused to answer (50%). Disagreement between the 2 methods was worst among the 33 individuals who described themselves with terms from 2 or more categories or who described themselves as multiracial (Table 4). Of these individuals, 18 (54.5%) chose a single race category, 9 (27.3%) chose “multiracial,” and 6 (18.2%) chose Other.

Time for Completion

The mean time required to collect patients’ self-described race/ethnicity with the computerized instrument tested in phase 2 was 37 ± 17 seconds. By comparison, the mean time for completing the 2 OMB questions was 20 ± 11 seconds.

DISCUSSION

This study shows it is possible to rapidly collect detailed information on patients’ race and ethnicity with open-ended questions that allow people to respond with terms of their choice. This allows maximal flexibility and adaptability. Health care providers can conduct analyses using the most specific categories possible, such as “How many Pakistani patients did we care for last year, and what proportion of Pakistani women received screening for cervical cancer?” This high level of detail will allow providers to identify vulnerable populations that should be targeted for quality-improvement efforts.²⁴ If an organization needs to report data to meet federal or state reporting requirements, providers can still aggregate unique terms to correspond to the major categories used by federal agencies and achieve similar results to what they would have obtained if they collected race/ethnicity information using the OMB categories. Allowing patients to state their race/ethnicity with their own terms took an average of only 37 seconds, so providers should be reassured that collecting this level of detail does not pose a major time or cost burden.

The most important advantage of using open-ended questions to collect information about patients’ race and ethnicity is improved accuracy. Many of the patients who identified themselves as Latino, Hispanic, or being from a country in Central or South America chose their race as Other or refused to answer the race question at all when they were presented

with the standard OMB question. This result is consistent with previous reports that many Latinos do not view themselves as having a race separate from their ethnicity.^{25,26} This reticence probably explains why Latinos/Hispanics in our study were much more likely to prefer to use their own words to describe their race and ethnicity (adjusted odds ratio = 3.0, 95% CI = 1.2, 7.5).

Similarly, individuals from the Middle East, the Indian subcontinent, Southeast Asia, and the Pacific Islands often do not describe themselves as White or Asian, and neither term is a meaningful indicator of their culture, their beliefs about health and health care, or their experiences within our health care system. As a consequence, many of the individuals from these groups in our study chose Other as their race or refused to answer the question. Using the OMB questions to collect race/ethnicity results in an unacceptably high rate of unusable data for individuals in these groups as well as Latinos/Hispanics, which could create bias in analyses of disparities in care and diminish the power to detect disparities among vulnerable groups.

Collecting race/ethnicity by allowing people to describe themselves with their own terms is particularly important for individuals who identify with 2 or more racial or ethnic groups. First, this system allows this group to identify themselves more precisely. Second, a majority (51.5%) of multiracial/multiethnic study participants said they somewhat or strongly preferred to describe themselves with their own terms as opposed to choosing from a list. However, we did not allow people to choose multiple racial categories from the OMB list. It is possible that the preference among multiracial individuals for describing themselves with their own terms would have been less if we had allowed people to choose multiple options from the list of race choices.

Despite these advantages, allowing patients to describe themselves with their own terms has limitations as well. Three patients described themselves only as Black or African American, but they said that they were Latino/Hispanic when asked this question directly. The fact that these patients did not describe themselves with the term *Latino* or *Hispanic* might mean that this aspect of their

heritage was not a very important part of their identity, although we did not specifically ask patients about this. Regardless, our results suggest that a small percentage of people with Latino/Hispanic ancestry may not be captured if patients describe their race/ethnicity with their own terms without being asked specifically whether or not they are Latino/Hispanic.

Limitations

There are several important limitations to our study. People who had negative attitudes toward answering questions about their race/ethnicity may have been less likely to participate. Therefore, our results may underestimate the proportion of people in the general population who would refuse to answer questions about their race/ethnicity. Although we could not obtain detailed information about reasons for refusals, the most common reason that patients spontaneously gave was that they did not have time. Eligible subjects who refused to participate were somewhat more likely to be White (57.5%) and less likely to be Hispanic (5.9%) than participants, but the proportion of women and the estimated age was similar to that of participants. Thus, the effect of nonresponse bias on our findings is likely to be relatively small.

We were also unable to interview people who did not speak English fluently. Approximately 3% of patients at our institution prefer to speak a language other than English, and approximately 1% have limited English proficiency. Most of these individuals are likely to have strong ethnic identities. Including them probably would have increased the number of people who used specific terms to describe their race/ethnicity and who chose Other or had missing data for the OMB race question. Although we did not interview patients with limited English proficiency or ask about language barriers as part of this study, routinely asking patients their preferred language and their ability to speak and understand English is probably just as important as collecting data on race/ethnicity. More research is needed to understand patients’ attitudes about collecting this information and the accuracy of their self-reported English proficiency.

We also do not know the generalizability of our findings. The study sample was

relatively small and had limited representation of Latinos, Asians, and other groups. The study was conducted at only 1 site, and results could vary across the country. Although the agreement between the OMB classification and patients' self-reported race/ethnicity was high in this study, the discrepancies found suggest that our agreement rate would have been markedly lower among a study population with a higher proportion of Latinos/Hispanics, people from the Middle East or South Asia, and multiracial/multiethnic individuals. Finally, we do not know health care providers' typical costs of implementing changes in data collection for patients' race/ethnicity. At Northwestern Memorial Hospital, it was possible to implement a new system with only a few days' effort by a computer programmer and several 1-hour sessions to train registration staff in new techniques. This likely varies significantly depending on the hardware and software used for registration. Training materials are now available over the Internet (available at: <http://www.hret disparities.org/hret disparities/html/general/gtraining.html>), so the cost of developing training materials is negligible.

Conclusions

Our previous work has shown that most patients think that health care providers should routinely collect race/ethnicity information and use it to monitor disparities in quality of care.²² The current study suggests that health care providers can rapidly and accurately accomplish this task with computerized systems that allow patients to use their own words. If future studies with larger and more diverse patient populations confirm our findings, this may become the preferred approach for collecting race/ethnicity data. National organizations and the scientific community will need to set standards for using these data to examine disparities. Standardized systems are needed to allow comparison of disparities across provider organizations and to allow providers to aggregate data from different sites to examine disparities for patients belonging to smaller, potentially vulnerable racial/ethnic groups, such as recent immigrants from the Middle East and South Asia. ■

About the Authors

David W. Baker, Kenzie A. Cameron, Joseph Feinglass, Jason Thompson, and Patricia Georgas are with the Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Ill. D. W. Baker is also with the Institute for Healthcare Studies, Feinberg School of Medicine. Shawn Foster, Deborah Pierce, and Romana Hasnain-Wynia are with the Health Research and Educational Trust, Chicago.

Requests for reprints should be sent to David W. Baker, MD, MPH, 676 N St Clair St, Suite 200, Chicago, IL 60611 (e-mail: dwbaker@northwestern.edu).

This article was accepted June 1, 2005.

Contributors

D.W. Baker, K.A. Cameron, J. Feinglass, and R. Hasnain-Wynia designed the study. P. Georgas, S. Foster, and D. Pierce interviewed patients. D.W. Baker, K.A. Cameron, and J.A. Thompson analyzed the data. D.W. Baker and K.A. Cameron prepared the article. All authors helped to design the survey and reviewed and revised the article.

Acknowledgment

This study was funded by The Commonwealth Fund and the Health Research and Educational Trust.

Human Participant Protection

The institutional review board for Northwestern University approved this project.

References

- Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine; 2003.
- The Right to Equal Treatment*. Washington, DC: Physicians for Human Rights; 2004.
- Fiscella K, Franks P, Gold MR, et al. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA*. 2000;283:2579–2584.
- Bierman AS, Lurie N, Collins KS, et al. Addressing racial and ethnic barriers to effective health care: the need for better data. *Health Aff (Millwood)*. 2002;21:91–102.
- Panel on DHHS Collection of Race and Ethnicity Data. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: National Academies Press; 2004.
- Hasnain-Wynia R, Pierce D, Pittman MA. *Who, What, When, Where: The Current State of Data of Collection on Race and Ethnicity in Hospitals*. New York, NY: Commonwealth Fund; 2004.
- Boehmer U, Kressin NR, Berlowitz DR, et al. Self-reported vs administrative race/ethnicity data and study results. *Am J Public Health*. 2002;92:1471–1472.
- Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in US infants. A new look at infant mortality, 1983 through 1985. *JAMA*. 1992;267:259–263.
- Swallen KC, Glaser SL, Stewart SL, et al. Accuracy of racial classification of Vietnamese patients in a population-based cancer registry. *Ethn Dis*. 1998;8:218–227.
- Swallen KC, West DW, Stewart SL, et al. Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Ann Epidemiol*. 1997;7:200–206.
- Pan CX, Glynn RJ, Mogun H, et al. Definition of race and ethnicity in older people in Medicare and Medicaid. *J Am Geriatr Soc*. 1999;47:730–733.
- Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Washington, DC: US Census Bureau. Available at: <http://www.census.gov/population/www/socdemo/race/Ombdir15.html>. Accessed January 16, 2005.
- Weinick RM, Jacobs EA, Stone LC, et al. Hispanic healthcare disparities: challenging the myth of a monolithic Hispanic population. *Med Care*. 2004;42:313–320.
- Hispanic health in the United States. Council on Scientific Affairs. *JAMA*. 1991;265:248–252.
- Trevino FM, Moyer ME, Valdez RB, et al. Health insurance coverage and utilization of health services by Mexican Americans, mainland Puerto Ricans, and Cuban Americans. *JAMA*. 1991;265:233–237.
- Mendoza FS, Ventura SJ, Valdez RB, et al. Selected measures of health status for Mexican-American, mainland Puerto Rican, and Cuban-American children. *JAMA*. 1991;265:227–232.
- Stroup-Benham CA, Trevino FM. Reproductive characteristics of Mexican-American, mainland Puerto Rican, and Cuban-American women: Data from the Hispanic Health and Nutrition Examination Survey. *JAMA*. 1991;265:222–226.
- Becerra JE, Hogue CJ, Atrash HK, et al. Infant mortality among Hispanics. A portrait of heterogeneity. *JAMA*. 1991;265:217–221.
- Zambrana RE, Carter-Pokras O. Health data issues for Hispanics: implications for public health research. *J Health Care Poor Underserved*. 2001;12(1):20–34.
- Gomez SL, Kelsey JL, Glaser SL, et al. Immigration and acculturation in relation to health and health-related risk factors among specific Asian subgroups in a health maintenance organization. *Am J Public Health*. 2004;94:1977–1984.
- Lee J, Heng D, Chia KS, et al. Risk factors and incident coronary heart disease in Chinese, Malay and Asian Indian males: the Singapore Cardiovascular Cohort Study. *Int J Epidemiol*. 2001;30:983–988.
- Baker DW, Cameron KA, Feinglass J, et al. Patients' attitudes toward health care providers collecting information about their race and ethnicity. *J Gen Intern Med*. 2005;20:895–900.
- Landis JR, Koch GG. The measure of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Greenfield A. New hospital gown gives Muslim women comfort [press release]. Maine Medical Center; June 22, 2004. Available at: http://www.mmc.org/mmc_news/062204.htm. Accessed January 16, 2005.
- Zubizarreta P. Is there a Hispanic race? *Lancet*. 1999;354:780.
- Navarro M. Going beyond Black and White, Hispanics choose "other." *New York Times*. November 9, 2003.